Article:
Summers, Robert J, Baker, Daniel Hart orcid.org/0000-0002-0161-443X and Meese, Tim S (2015) Area summation of first- and second-order modulations of luminance. Journal of Vision. 12. pp. 1-13. ISSN 1534-7362

https://doi.org/10.1167/15.1.12

Area summation of first- and second-order modulations of luminance

Robert J. Summers
Centre for Vision and Hearing Research, School of Life and Health Sciences, Aston University, Birmingham, UK

Daniel H. Baker
Department of Psychology, University of York, Heslington, York, UK

Tim S. Meese
Centre for Vision and Hearing Research, School of Life and Health Sciences, Aston University, Birmingham, UK

To extend our understanding of the early visual hierarchy, we investigated the long-range integration of first- and second-order signals in spatial vision. In our first experiment we performed a conventional area summation experiment where we varied the diameter of (a) luminance-modulated (LM) noise and (b) contrast-modulated (CM) noise. Results from the LM condition replicated previous findings with sine-wave gratings in the absence of noise, consistent with long-range integration of signal contrast over space. For CM, the summation function was much shallower than for LM suggesting, at first glance, that the signal integration process was spatially less extensive than for LM. However, an alternative possibility was that the high spatial frequency noise carrier for the CM signal was attenuated by peripheral retina (or cortex), thereby impeding our ability to observe area summation of CM in the conventional way. To test this, we developed the "Swiss cheese" stimulus of Meese and Summers (2007) in which signal area can be varied without changing the stimulus diameter, providing some protection against inhomogeneity of the retinal field. Using this technique and a two-component subthreshold summation paradigm we found that (a) CM is spatially integrated over at least five stimulus cycles (possibly more), (b) spatial integration follows square-law signal transduction for both LM and CM and (c) the summing device integrates over spatially-interdigitated LM and CM signals when they are co-oriented, but not when cross-oriented. The spatial pooling mechanism that we have identified would be a good candidate component for a module involved in representing visual textures, including their spatial extent.

Introduction

First- and second-order vision

The human visual system is sensitive to both luminance-modulated (LM) and contrast-modulated (CM) stimuli and has mechanisms well suited to the extraction of each of these sources of information from the retinal image (e.g., Foley & Legge, 1981; Malik & Perona, 1990). The contrast detection and discrimination of simple first-order stimuli (e.g., sinusoidal luminance gratings) is typically modeled by a linear spatial filter followed by a sigmoidal transducer that accelerates for low stimulus contrast (e.g., a square-law; Meese & Summers, 2009; Meese, 2010) but becomes compressive at higher contrasts (Legge & Foley, 1980). The processing of second-order, or non-Fourier stimuli—so-called because the contrast-modulation component is not directly visible in the Fourier spectrum (e.g., Schofield, 2000)—is characterized by a filter-rectify-filter (FRF) model (e.g., Dakin & Mareschal, 2000; Schofield, 2000) in which a band-limited filter is followed by a rectifying nonlinearity then a second-stage band-limited filter tuned to lower spatial-frequencies than the first stage. The FRF models of second-order vision were influenced by related models of texture segmentation (e.g., Malik & Perona, 1990; Landy & Bergen, 1991) and second-order motion perception (e.g., Chubb & Sperling, 1988; Wilson, Ferrera, & Yo, 1992).

The extent to which first- and second-order stimuli are processed independently is a matter of some debate.

Citation: Summers, R. J., Baker, D. H., & Meese, T. S. (2015). Area summation of first- and second-order modulations of luminance. *Journal of Vision, 15*(1):12, 1–13, http://www.journalofvision.org/content/15/1/12, doi:10/1167.15.1.12.

doi: 10.1167/15.1.12 Received October 10, 2014; published January 14, 2015 ISSN 1534-7362

There is no facilitation from a subthreshold pedestal between LM and CM, either way (i.e., there is no "dipper;" Schofield & Georgeson, 1999), nor is there masking between these cues (Allard & Faubert, 2007); they are discriminable at detection threshold (Georgeson & Schofield, 2002) and the judgment of global orientation using one cue is unaffected by the presence of the other (Allen, Hess, Mansouri, & Dakin, 2003). All of this points to independent channels for LM and CM. At some stage, however, CM and LM information must be integrated. For example there is energy summation between superimposed co-oriented LM and CM gratings (Schofield & Georgeson, 1999) and orientation and contrast after-effects transfer from one cue to the other (Georgeson & Schofield, 2002).

Area summation in first- and second-order vision

As a luminance grating is increased in size its contrast detection threshold decreases. When this is done for a grating placed in the central visual field, the initial benefit of area soon diminishes, such that detection thresholds improve very little beyond eight signal cycles or so. The asymptotic nature of the area summation function is traditionally attributed to retinal inhomogeneity (e.g., see Baldwin, Meese, & Baker, 2012; Bradley, Abrams & Geisler, 2014) and, until recently, the benefit of stimulus area was thought to be the result of probability summation over the relevant detecting mechanisms (Robson & Graham, 1981; Pelli, 1985; Tyler & Chen, 2000). However, by manipulating extrinsic uncertainty and appealing to constraints from the shape of the psychometric function, and by measuring classification images, a series of recent papers (Meese & Summers, 2007, 2009, 2012; Meese, 2010; Baker & Meese, 2011, 2014; Baldwin & Meese, in preparation) has provided strong evidence that probability summation cannot account for the area summation effects in first-order modulations of luminance. Instead, the results are consistent with a noisy energy model (Meese, 2010; Meese & Summers, 2012) that sums the squared outputs of noisy linear filter-elements over space. Thus, the contrast detection of grating-like stimuli reveals linear spatial pooling over much larger areas ($\geq$ 8 cycles) than has previously been supposed.

As alluded to above, one problem with the traditional area summation experiment is that nonuniform contrast sensitivity across the visual field means that the more peripheral parts of the stimulus contribute little, if at all, to performance because the detecting mechanisms there are so much less sensitive than those in the central fovea. In fact, beyond a critical distance, peripheral mechanisms would degrade the signal-to-noise ratio (SNR) by contributing disproportionate amounts of noise (Meese & Summers, 2012). One way to address this problem is to keep the stimulus diameter constant while manipulating the area of the stimulus that contains the signal. Using novel interdigitated modulation of signals in so-called Swiss cheese stimuli (Meese & Summers, 2007, 2009; Meese & Baker, 2011) and Battenberg stimuli (Meese, 2010), summation ratios of ~6 dB (a factor of 2) have been found for a doubling of stimulus area. This contrasts with the results from traditional area summation experiments (c.f. Meese & Summers, 2012), where the same improvement in performance (~6 dB) requires the stimulus area to be quadrupled by increasing grating diameter from one to two cycles. In extensive testing of several types of summation model, including a contemporary implementation of probability summation (by taking the max() over detecting mechanisms; see Tyler & Chen, 2000), only the noisy energy model was found to be consistent with the full set of results (Meese & Summers, 2012).

Studies of spatial summation in second-order vision have found similar bow-shaped summation curves for second-order stimuli such as CM Gaussian blobs (Schofield & Georgeson, 1999; Sukumar & Waugh, 2007), CM gratings (Wong & Levi, 2005). and narrow-band orientation-filtered noise textures (Landy & Oruç, 2002). In those studies that compared spatial summation of CM and LM stimuli, both Schofield and Georgeson (1999) and Wong and Levi (2005) found that the summation curves for CM and LM were parallel across the whole range of stimuli tested. However, Sukumar and Waugh (2007) found that the area summation for CM extended beyond the range of LM. Notwithstanding the differences above, the similarity of the LM and CM functions implies that the noisy energy model might be extended to also account for area summation of CM stimuli.

More generally, it is also natural to inquire whether the Swiss cheese approach developed above can be exploited to shed new light on area summation for second-order (CM) signals. In particular, we wanted to know whether the first-order area summation process that we have identified previously (see above) also receives input from second-order stimuli. A result of that kind would suggest that the area summation process is a fairly general purpose mechanism, operating on a variety of inputs, as might be appropriate for marking visual textures, for example.

The experiments presented here compared area summation in CM and LM stimuli when: (a) signal diameter and area increased together (as in the traditional method), and (b) signal diameter was held constant but signal area was increased by filling in the holes of a "Swiss cheese" (Meese & Summers, 2007, 2009). To investigate the generality of spatial pooling, a

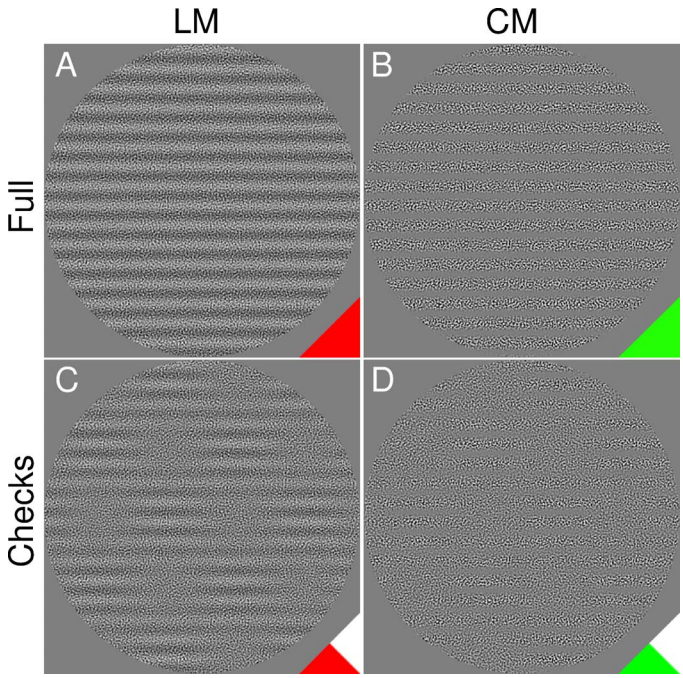

Figure 1. Examples of 16-cycle single-component stimuli for Experiment 1. A and B are CM and LM noise gratings with a full envelope. C and D are equivalent, but with a checks envelope. Note that in all cases, the 1.25 c/° grating modulation was the signal.

second experiment compared summation between CM- and LM-signals, where the modulations of the two components in the compound were either superimposed, interdigitated or cross-oriented.

Methods

Equipment

Stimuli were presented on a Sony Multiscan 20 SE-II CRT (Sony, Tokyo, Japan) using a ViSaGe (CRS Ltd, Cambridge, UK) operating in 14-bit mono mode. The display was linearized (Berns, Motta, & Gorzynski, 1993) and calibration was performed daily. Resolution of the display was 1024 × 768 at a frame-rate of 100 Hz with mean luminance of 53 cd/m^2. Observers sat in a dark room at a viewing distance of 1 m with their head in a chin and head rest, and viewed the stimuli binocularly. At this distance, 55 pixels subtended one degree of visual angle.

Stimuli

Contrast- and luminance-modulated noise stimuli were created with the following equation

$$\mathbf{S} = L\times[1 + \mathbf{N} + (c_{CM} \times \mathbf{N} \times \mathbf{G_{CM}} \times \mathbf{E_{CM}}) + (c_{LM} \times \mathbf{G_{LM}} \times \mathbf{E_{LM}})] \quad (1)$$

where L is mean luminance, $\mathbf{N}$ is a two-dimensional sample of band-limited zero mean noise, $\mathbf{G}_{CM}$ and $\mathbf{G}_{LM}$ are sine-wave gratings, $\mathbf{E}_{CM}$ and $\mathbf{E}_{LM}$ are the grating envelopes, and c_{CM} and c_{LM} are the modulation contrasts of the CM and LM signal components. LM-only stimuli were created by setting c_{CM} to zero. CM-only stimuli were created by setting c_{LM} to zero.

The target gratings ($\mathbf{G}_{CM}$ and $\mathbf{G}_{LM}$) were all sinusoidal (spatial frequency [SF] = 1.25 c/°; 44 pixels per cycle) and were modulated by a circular window with a central plateau of unity, and half a cycle of a raised-cosine skirt (2 pixels wide) around the edge. The nominal grating diameters were 1, 2, 4, 8, and 16 cycles (12.8°). Since the raised-cosine skirt extended beyond the central plateau of the grating, the smallest stimulus had a diameter of 48 pixels (44 for the one cycle grating plus 2 either side for the modulator) and the largest had a diameter of 708 pixels. In Experiment 1 both $\mathbf{G}_{CM}$ and $\mathbf{G}_{LM}$ were oriented horizontally. In two conditions of Experiment 2, $\mathbf{G}_{LM}$ was oriented vertically. The gratings were always in sine-phase with respect to the stimulus center such that the peak was always either below or to the right of fixation depending on the orientation of the grating.

Further, there were two conditions for the shape of the overall stimulus envelope. The first was a "full" (i.e., uniform) condition and was exactly as described above. The second was a "checks" condition, in which a raised plaid modulator (Meese & Summers, 2007) was applied to the window described above. The check spatial frequency was 0.25 c/° (horizontal and vertical). There were 2.5 grating cycles per modulator check. Interdigitated LM/CM compounds were created by setting $\mathbf{E}_{LM}$ in counter-phase with $\mathbf{E}_{CM}$. This check modulation was used only for the eight- and 16-cycle stimuli. The phase of the checks envelope was fixed (see Figure 1).

Note that the full stimulus can be thought of as a compound stimulus constructed by summing (a) the signal in a checks stimulus with (b) the signal in another checks stimulus of opposite phase. Thus, by measuring sensitivities to full and check stimuli we were performing a classical two-component summation experiment.

The stimulus noise, $\mathbf{N}$, was bandpass-filtered Gaussian white noise (center SF = 8 c/°; ±0.5 octave bandwidth) with RMS contrast of 20%. The noise always subtended 708 pixels (which was the diameter of the largest signal). The filtering was necessary for two reasons. First, to ensure the modulation frequency and carrier frequency were spectrally distinct, thereby avoiding within-channel masking effects (Schofield & Georgeson, 1999); second, to avoid "clumping" arte-

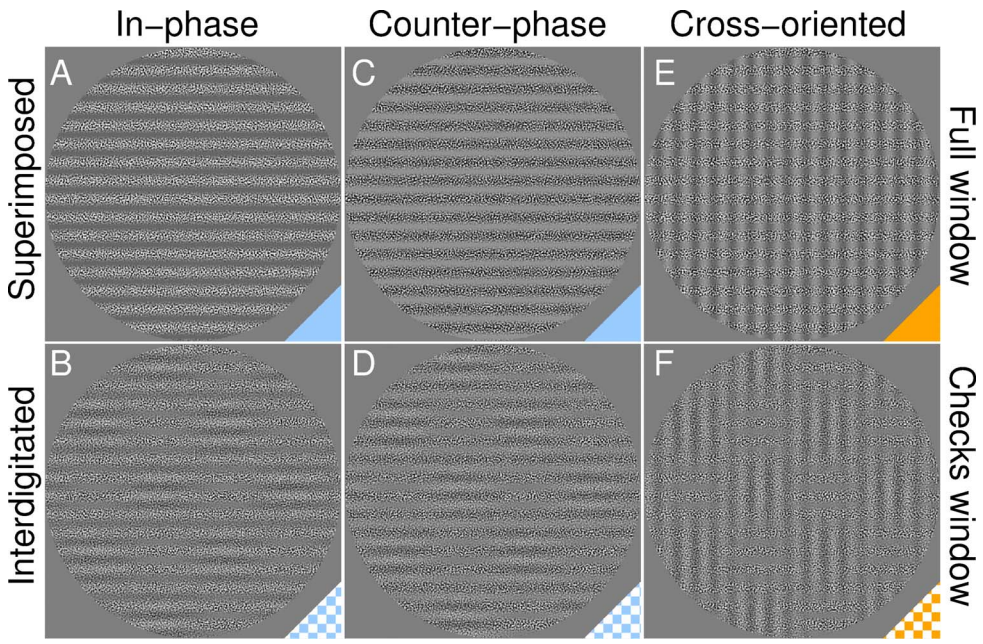

Figure 2. Two-component stimuli for Experiment 2. A and C are mixtures of horizontal CM and LM gratings with a full envelope where the gratings are (A) in phase and (C) in counter phase with each other. B and D are the equivalent interdigitated conditions; each component (LM and CM) has a checks envelope in counter-phase with the other. E is a cross-oriented CM/LM mixture where the LM component is a vertical grating and there is a full envelope for each grating. F is similar to E, but each component has a checks envelope as in B and D. In the experiment the relative levels of each component were normalized for each observer with respect to their sensitivity. Note that in all cases, the 1.25 c/° grating modulation was the signal.

facts (Smith & Ledgeway, 1997) that might cause second-order stimuli to be detected by first-order mechanisms. In addition, the possibility of side-band detection of CM stimuli by LM mechanisms (Dakin & Mareschal, 2000) is avoided by the use of an isotropic noise carrier.

Fixation marks indicating the spatial extent of the target were displayed throughout the experiment. They comprised a quad of points (2 × 2 pixels) at the corners of a virtual square that enclosed the signal modulation (the length of one side of this square was equal to the diameter of the grating, $\mathbf{G}_{CM}$ or $\mathbf{G}_{LM}$). We did not use a central fixation point as this can interfere with the detection of small stimuli (Meese & Hess, 2007; Summers & Meese, 2009).

Examples of the single- and two-component stimuli are shown in Figures 1 and 2 respectively.

Procedure

Target modulation contrast was selected by a pair of interleaved three-down/one-up staircases (Wetherill & Levitt, 1965). Step-size was 3 dB and each staircase terminated after 12 reversals (~100 trials for the pair). Thresholds (at 81.6% correct) were obtained by fitting a two free-parameter (threshold and slope) Weibull function to the staircase data using *psignifit* (Wichmann & Hill, 2001). Mean observer thresholds for each stimulus configuration were derived by averaging the fitted threshold parameter estimated across four blocks (~400 trials).

We used a temporal two-interval forced-choice (2IFC) procedure where, from trial to trial, one randomly selected interval contained a nonmodulated noise stimulus and the other contained a modulated-noise stimulus (the signal) at the modulation depth selected by the staircase. The noise sample in each interval was always different. Stimulus duration was 100 ms and the interstimulus interval was 400 ms. Each interval was accompanied by a tone to reduce temporal uncertainty. Observers selected the interval in which they perceived the signal using the buttons of a mouse. Auditory feedback was provided to indicate correctness of response.

In Experiment 1, each experimental session comprised threshold estimates for either CM or LM alone, blocked by stimulus diameter and envelope (full or checks). In Experiment 2, the single-component (LM or CM) thresholds obtained in Experiment 1 were used to create CM+LM mixtures where the component magnitudes were adjusted to normalize sensitivities on an observer by observer basis. After completing Experiment 1, we also performed a supplementary experiment in which we measured detection thresholds for lumi-

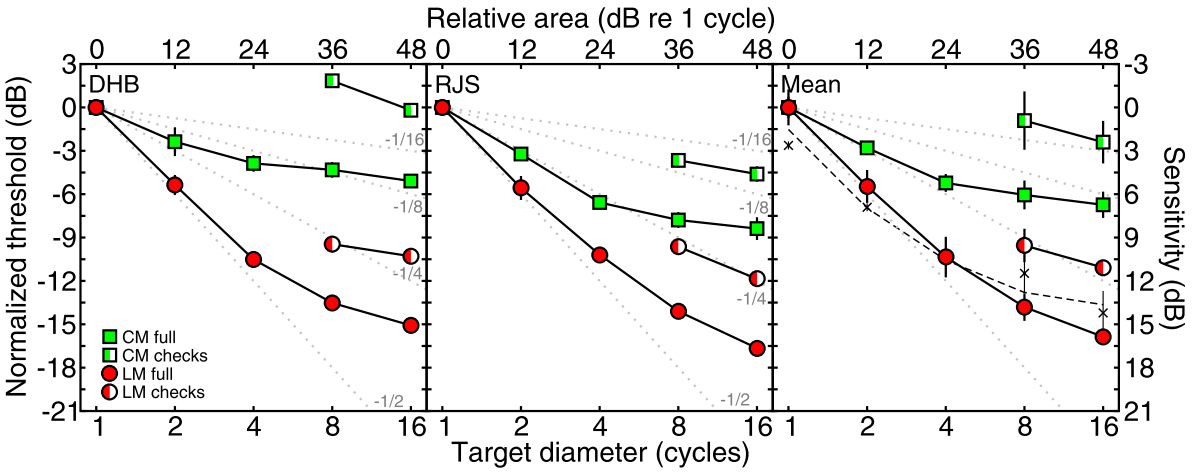

Figure 3. Results for Experiment 1: area summation functions. Contrast detection thresholds are for DHB (left), RJS (middle) and their average (right), for LM signals (circles) and CM signals (squares). Solid and half-filled symbols are for signals with a full (i.e., uniform) and a checks envelope, respectively. Thresholds for full CM and full LM signals are normalized to 0 dB for a one-cycle image. The differences between CM and LM thresholds for one-cycle targets were 22.1 dB and 21.0 dB for DHB and RJS, respectively. The results for the checks stimuli are expressed relative to their respective full conditions. Error bars are standard errors of the within-subject mean (left and middle panels) and the between-subjects mean (right panel). In the right hand panel the dashed curve is the prediction of the noisy energy model (no free parameters; see text for details) and the black crosses are mean thresholds for luminance gratings (the results of the supplementary experiment for DHB and RJS). These are normalized so that their average is matched to that of the mean LM full results (see text for details). In all cases, the signal spatial frequency was 1.25 c/°.

nance gratings in the absence of noise (i.e., **N** in Equation 1 was set to zero.)

Observers

Three experienced observers took part in the experiments. DHB and RJS (co-authors) completed Experiments 1, 2, and the supplementary experiment. SAW completed Experiment 2 and the subset of conditions from Experiment 1 that were needed for Experiment 2. SAW was naive to the purposes of the experiment. Observers had normal or optically corrected-to-normal vision.

Results and discussion

Experiment 1: A comparison of area summation of LM and CM signals

The detection thresholds for 1.25 c/° LM and CM gratings (one to 16 cycles; full/uniform window) are shown in Figure 3. For LM signals (filled circles) the summation curves are bow-shaped, similar to those found for sine-wave gratings in the absence of noise, and shown by the cross symbols in Figure 3C (the results of the supplementary experiment). The normalization procedure hides the overall difference between these two functions, but overall sensitivity was lower in the presence of external noise (e.g., a difference of 5.3 dB at one cycle), owing to cross-channel suppression from the bandpass noise (Baker & Meese, 2012). This confirms that area summation of luminance contrast remains intact in the presence of noise masks, consistent with previous observations using cross-oriented masks (Meese, 2004).

We now compare predictions from the zero-free-parameter noisy energy model (Meese, 2010; Meese & Summers, 2012)—see black dashed curve on the right-most panel of Figure 3—with the area summation results for our LM stimuli. Full details of this model are given in Meese and Summers (2012); we provide only an overview here. First, the input image was filtered (by a 1.25 c/° horizontal sine-phase filter: spatial-frequency bandwidth of 1.6 octaves, orientation bandwidth of ±25°; Meese, 2010) and multiplied by a retinal attenuation surface to simulate spatial inhomogeneity (using the average parameters from Table 5 in Baldwin et al., 2012). Each pixel was then squared and summed over area. The prediction is fair, but slightly underestimates the overall effects of summation. For comparison, the crosses in Figure 3C are the mean and standard error for sensitivity to a 1.25 c/° luminance grating (with no noise carrier). The model predicts these results very well, suggesting that the minor deviations of the model from the LM results owes to a subtle effect of noise masking, beyond the scope of our modeling here.

An experiment by Wong and Levi (2005) is closely related to our Experiment 1, but there are several differences in the results (and methodologies) across the two studies. We think that our use of narrow-band noise gave us the cleaner measure of the processes being investigated, but some of the differences across the two studies remain puzzling. We shall consider these differences in the General Discussion. However, one puzzle that we shall address here is the question of why our CM summation curve is so much shallower than our LM summation curve. Assuming that area summation of CM signals is determined by linear spatial summation across noisy local CM mechanisms, there are at least three possibilities: (a) spatial summation is less extensive for CM than for LM, (b) the overall transducer exponent (p) for the local CM mechanism (before area summation) is much greater than for LM (i.e., $p \gg 2$; c.f. figure B3 in Meese & Summers, 2012), or (c) CM sensitivity is attenuated more heavily with eccentricity than is LM sensitivity. We shall present arguments against the first two possibilities later. The third possibility is more promising. For example, Hess, Baker, May, and Wang (2008) demonstrated that for low-modulation frequencies (~1 c/° and below), the decline in CM sensitivity with eccentricity follows that predicted by the spatial-frequency of the carrier. As the deleterious eccentricity effects for LM signals are scale invariant (e.g., Baldwin et al., 2012), and as the center spatial-frequency of our band-limited noise carrier was 8 c/°, 6.4 times higher than our 1.25 c/° signals, we might expect signal sensitivity to decline with eccentricity much more rapidly for CM than for LM, consistent with our results (Figure 3). Notwithstanding our remark above regarding the choice of noise carrier, the considerations here reveal a profound difficulty in deriving a true measure of spatial summation of second-order signals using conventional methods.

We addressed the problem above by extending Experiment 1 to use the method of Meese and Summers (2007), where contrast sensitivity is compared for stimuli with a fixed diameter—thereby stimulating the same retinal extent—but a variable signal area. The half-filled symbols in Figure 3 show detection thresholds for check windowed stimuli (for LM and for CM; see Figure 1C and D) with diameters of eight and 16 cycles.

For both LM and CM stimuli, filling in the gaps of the checks to produce full (uniform) modulation—thereby doubling the signal area—improved sensitivity by an average of about 4.7 dB (a factor of 1.7; compare filled and half-filled symbols in Figure 3). Clearly there was substantial area summation for both CM and LM stimuli (over at least two envelope checks, equivalent to five stimulus cycles) despite the relatively small improvement in sensitivity when area was increased in the conventional way (by increasing stimulus diameter). The "cheese effect" here is similar in magnitude to the ~5 dB benefit that has been found for luminance gratings with comparable windowing (Meese & Summers, 2007; Baker & Meese, 2011).

Experiment 2: Investigating the combination of LM and CM over space

Experiment 1 provided good evidence for extensive area summation of signal for both LM and CM signals (well beyond that of a stereotypical V1-type local receptive field). However, it remained unclear whether there is a single summing device for both signal types, or parallel pathways for area summation across LM and CM. Previous studies (Schofield & Georgeson, 1999) have reported results consistent with energy summation between superimposed LM and CM gratings of the same orientation, but did not establish whether this was taking place only within local mechanisms or whether a more global signal integrator might be involved. Here we addressed this issue with the help of the check windowed stimulus design: by placing patches of CM and LM signals on either the same or different parts of the retina, it becomes possible to determine whether the area summation device receives input from both types of signal. This follows the approach of Meese and Summers (2009), who used the technique to investigate contrast (LM) summation across area and eyes.

Specifically, we compared summation ratios for a single component (either LM or CM) and several CM + LM compounds, where the CM and LM components were either superimposed or interdigitated across space. Furthermore the two components had either the same or orthogonal orientations (co-oriented or cross-oriented). When the stimulus components were co-oriented, the signals were either in- or out-of-phase with each other. For the compound stimuli and for each observer the signal strength of the CM component was adjusted relative to the LM component so that sensitivity was equalized across the two. For the co-oriented component pairs, this normalization was done using the mean thresholds from Experiment 1. For the cross-oriented conditions normalization was carried out relative to thresholds obtained for full and check windowed vertical LM stimuli in Experiment 2. In all cases, the normalization was done on an observer by observer basis, and stimulus diameter was 16 cycles.

For each stimulus pairing, summation ratios were calculated as the dB difference (i.e., 20 × log of the sensitivity ratio) for a single component in the compound stimulus (LM + CM) compared to that single component alone (the normalization procedure means that it does not matter which). The results for

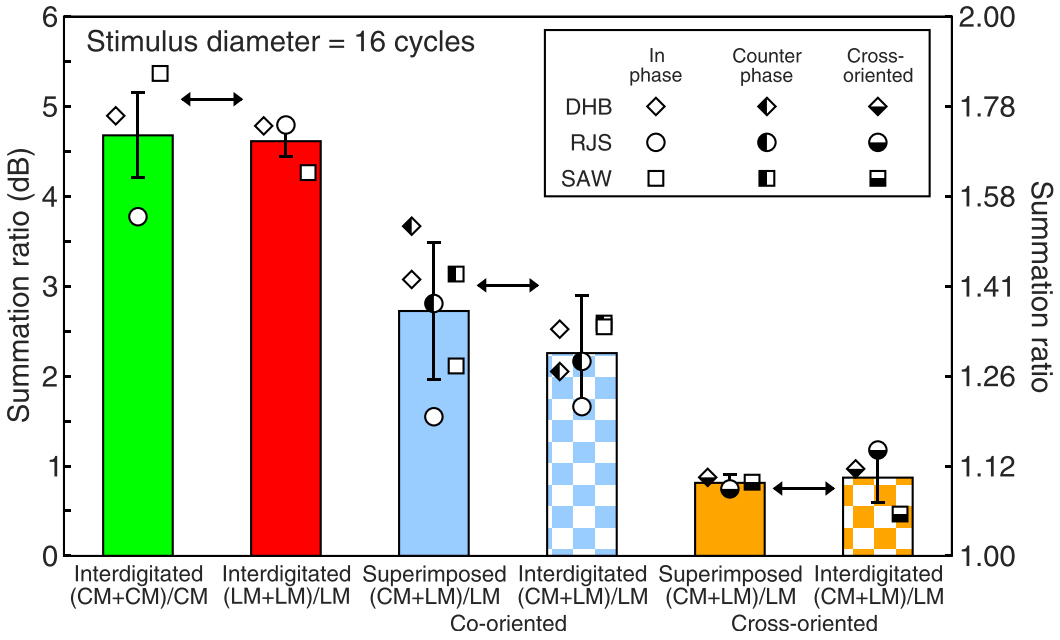

Figure 4. Results for Experiment 2: Summation ratios are for various pairs of stimulus components. Red and Green: Benefit of the full (uniform) windowing over the check windowing for CM and LM signals, respectively (see Figure 1). Blue, Solid, and Hashed: Benefit of adding a normalized CM signal to an LM signal for superimposed and interdigitated component pairs, respectively. Results were averaged across phase (see Figure 2A–D). Orange, Solid, and Hashed: same as blue, but for crossed signal orientations (see Figure 2E, F). Bars show means across three observers and interobserver standard errors; symbols are individual means for each condition. The double-headed arrows are predictions from a zero free-parameter model for each pair of conditions (see text for details). Note that the *x*-axis **/* labels refer to sensitivities (not thresholds).

three observers (symbols) and their average (bars), are shown in Figure 4. The two bars on the far left show the mean summation ratio between full and checks windowed stimuli for CM (green) and LM (red) signals (results for RJS and DHB are replotted from Experiment 1), which was 4.7 dB in each case. This repeats the observation from Experiment 1: signal is integrated over space for both first- and second-order stimuli.

The middle two bars show the results for co-orientation summation between CM and LM. When the signals were superimposed (solid blue bar) mean summation was 2.7 dB. This is close to the energy summation prediction (3 dB) found previously by Schofield and Georgeson (1999), and provides supporting evidence for a spatial mechanism that integrates signal over first- and second-order stimuli. Summation ratios were only slightly lower (2.3 dB; hashed blue bar), when the two components were interdigitated, suggesting that the mechanism first identified by Schofield and Georgeson also integrates over space. When the signals were superimposed (solid blue bar) there was a slight tendency for the counter-phase arrangement to produce higher summation ratios than the in-phase arrangement (3.2 dB vs. 2.2 dB on average; not shown). Previous work (Schofield & Georgeson, 1999; Georgeson & Schofield, 2002) found no effect of phase but, like us, Schofield and Georgeson found more interobserver variability in the in-phase condition than the out-of-phase condition (see their figure 7). It is not clear what the cause of these small differences and/or individual differences might be, but they detract little from our overall findings and conclusions.

The two bars on the far right show the summation ratios between orthogonal LM and CM signals. There was very little summation (0.9 dB, consistent with probability summation; see Modeling below) for these conditions, regardless of whether the components were superimposed (solid orange bar) or interdigitated (checked orange bar), less than has been found for interdigitated pairings of orthogonally-oriented LM signals (Meese, 2010). Note that this loss of cross-oriented summation relative to the co-oriented LM+CM conditions strengthens the case for there being something more sophisticated than probability summation in the co-oriented conditions, since otherwise, surely, we would expect a similar result for the cross-oriented conditions.

Modeling area summation of LM and CM

Schofield and Georgeson (1999) found that a simple two-channel energy summation model fit their data

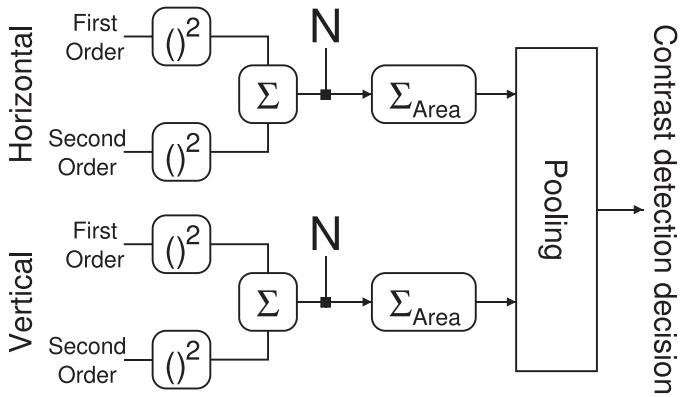

Figure 5. Area summation model for combining co-oriented signals across first- and second-order mechanisms and across space. Following Meese (2010), limiting noise (N) is placed before the area summation stage. For simplicity, summation across first- and second-order channels is shown as mandatory, but this is not a requirement of our model and a scheme in which the observer is able to switch between first- and second-order mechanisms as appropriate is also consistent with our results (i.e., the summing box to the left of the noise (N) has the option to switch as well as to sum), since this does not affect the limiting noise (N). Cross-oriented signals are assumed to be combined by probability summation at the "pooling" stage, following the area summation stage. This scheme extends the area summation scheme proposed by Meese (2010; his figure 6) for LM stimuli.

well. Their model computed the mean energy in each channel (i.e., the mean of the squared outputs of the LM and CM mechanisms, prior to linear summation). Here we extend the scope of that model by computing the LM and CM energy at each location in the image (i.e., summing the squared outputs of the LM and CM mechanisms at each location) prior to linear summation across space (see Figure 5). We develop this model first, producing the successful predictions in Figure 4, before considering and rejecting some alternatives.

We begin by making the simplifying assumption that our LM and CM mechanisms are able to successfully recover the signals from the stimulus (e.g., as shown by Schofield & Georgeson, 1999), and that they have equal responses at detection threshold. This allows us to use normalized signal strengths (c) as the inputs to a computational model. For co-oriented signals, the model involves linear summation of signal over space. Note that we have made no explicit provision for spatial phase but, led by our results (Figure 4), we assume that phase information is discarded at the initial filtering stage. Cross-oriented signals produced very little summation in our Experiment 2, and we wondered whether this might be accommodated by probability summation across cross-oriented mechanisms. Detail modeling of this concept can be quite complex (Tyler & Chen, 2000; Meese & Summers, 2012), but for the situation here it is reasonable to approximate these effects using Minkowski summation with an exponent of 4 across orthogonally-oriented mechanisms following their square-law nonlinearities (see Meese & Summers [2012] for extensive theoretical treatment of probability summation and data).

Following Meese (2010; appendix C), we derived an analytic expression for summation ratios between single- and dual-component stimuli. The signal-to-noise ratio (SNR) is given by:

$$\mathrm{SNR} = \frac{\sum |c \times \mathrm{LM}_i|^2 + |c \times \mathrm{CM}_i|^2}{\sqrt{n}} \quad (2)$$

Where LM_i and CM_i are co-oriented luminance- and contrast-modulated signals, respectively, at the ith spatial location, n is the number of pixels in the input image and $\sqrt{n}$ is the expected standard deviation of the noise in the summation channel. With the arbitrary assumption that signal detection occurs for signal strength, $c_{T,}$ (contrast at threshold) when $\mathrm{SNR} = 1$ then:

$$1 = \frac{c_T^2 \sum |\mathrm{LM}_i|^2 + |\mathrm{CM}_i|^2}{\sqrt{n}} \quad (3)$$

and solving for c_T we have:

$$c_T = \left[\frac{\sum |\mathrm{LM}_i|^2 + |\mathrm{CM}_i|^2}{\sqrt{n}}\right]^{-\frac{1}{2}} \quad (4)$$

The summation ratio, SR, for a co-oriented summation channel is given by:

$$\mathrm{SR} = \frac{c_T(\mathrm{Single})}{c_T(\mathrm{Compound})} \quad (5)$$

where c_T(Compound) and c_T(Single) are the signal strengths at threshold for the same single component in the compound and single-component stimuli respectively (because of the normalization it does not matter which).

Combination across orientation channels is assumed to be probability summation, following area summation, and is computed here using Minkowski summation with an exponent of 4 to combine contrast sensitivities, giving:

$$c_T(\mathit{CrossOrientedCompound}) = \left[\sum (c_{T,\theta=0^\circ}^{-4} + c_{T,\theta=90^\circ}^{-4})\right]^{-\frac{1}{4}} \quad (6)$$

where $c_{T,\theta=y}$ is the contrast detection threshold for the channel oriented at y^0. In this instance, the summation ratio is computed using Equation 5, but with Equation 6 substituted into the denominator.

Quantitative model predictions are shown in Figure 4 by the double-headed arrows. Note that the model

predicts exactly the same levels of summation for each condition in the pair, but different levels for each of the three pairs. In all cases, the predictions are in good agreement with the experimental results with no free parameters. When the two stimulus components are the same orientation and of the same type (i.e., both LM or both CM) then high summation (5.1 dB) is predicted and found (red and green bars in Figure 4). At first glance this might seem puzzling since in the model, linear summation over area comes after the square-law (energy) transduction, and so we might expect the conventional prediction of 3 dB (a factor of $\sqrt{2}$) for two-component energy summation. However, as recognized elsewhere (Meese & Summers, 2007, 2009), the raised antiphase pair of check modulators that sum to produce our full stimulus (see Methods) are not spatially independent but have substantial overlap in the regions between their peaks. This means that the two modulators sum locally in the display (prior to exponentiation by the square-law transducer in the model), lifting the summation ratio predicted by the model above 3 dB.

When the two components are co-oriented but of different types (LM + CM), then intermediate energy summation (3 dB) is predicted and found. Note that for this stimulus, although the LM and CM signals are independent; there is no physical summation between them in the display and the model follows the factor of $\sqrt{2}$ intuition. Note also that the global (area) integration property of the model means that predictions are identical, regardless of whether the LM and CM components were superimposed or interdigitated.

Probability summation for cross-oriented components, following energy transduction and area summation, predicts a summation ratio of 0.75 dB (the combination of square-law signal transduction and a Minkowski exponent of 4 produces an overall summation exponent of 8), and is very close to what we found for the cross-oriented pairings (0.81 dB and 0.9 dB for interdigitated and superimposed cross-oriented components respectively). The other conditions produced more summation than this, implying that they involved a summation process more potent than probability summation, consistent with our model.

We can also rule out several other model configurations. One possibility is that square-law energy transduction follows linear summation across LM and CM. However, if this were the case, our interdigitated LM + CM stimuli would behave in the same way as LM and CM alone (red and green bars in Figure 5)—linear summation in the visual system mimicking linear summation of signal in the display—and that is not what we found. Another possibility is that area summation takes place within LM and within CM channels before summing across the two. In this case, independent noise sources would need to be placed after square law transduction, one for each channel (LM and CM), but before the area summation device to be consistent with previous work (Meese, 2010; Meese & Summers, 2012). For this arrangement to predict 3 dB of summation between superimposed LM and CM stimuli (Figure 4) mandatory summation across LM and CM channels would be required. Without this, the observer could benefit by switching out the noise from the irrelevant channel in the single component conditions. This cascade of ideal summation with square-law energy transduction predicts a fourth-root summation law (Meese, 2010), giving 1.5 dB of summation for interdigitated LM + CM pairs, which is too little (by comparison to human results). Thus, the requirement for mandatory pooling across LM and CM channels (to achieve summation = 3 dB) makes this alternative scheme seem unlikely since LM and CM signals are clearly distinguishable at threshold (Georgeson & Schofield, 2002) and above (e.g., see Figure 1).

Summary and general discussion

We assessed area summation of luminance and contrast modulations of a noise stimulus. Summation slopes were shallower for CM than for LM signals (discussed below). However, when stimulus diameter was held constant and signal area was doubled, summation was around a factor of 1.7 within LM or CM stimuli. Some of this we attribute to signal summation within the display, the rest (a factor of $\sqrt{2}$) to energy summation over space. Combining the two cues either in a superimposed or interdigitated configuration reduced signal summation to levels roughly consistent with energy summation (a factor of around $\sqrt{2}$). Very little signal summation was found when LM and CM gratings were orthogonal. All the results are shown to be consistent with a noisy energy model that sums co-oriented LM and CM channels then pools over area, with a final stage of probability summation for cross-oriented signals. There is good evidence that co-oriented first- and second-order information is partially correlated in natural images (Johnson & Baker, 2004) and so it is perhaps not surprising that the visual system exploits this correlation by pooling across these signals. Our experiments here do not put an upper limit on the spatial extent of the pooling, but it must operate over at least five signal cycles.

Comparison with Wong and Levi (2005)

Wong and Levi (2005) found steeper summation functions than we did. In particular, for both LM and

CM they found the initial slope (from one to two cycles) to be −1. They attributed this to linear summation within a receptive field, but this is questionable. A slope of −1 would occur only if the limiting noise were constant (see Meese, 2010). While this is true for their full-field external noise, the implication is that the same detecting mechanism was used when the number of stimulus cycles was doubled from one to two cycles (thereby doubling the integral of signal contrast but with no concomitant increase in noise). In the blocked experimental design used by Wong and Levi (2005) (and us) this seems unlikely, particularly for LM, since most estimates put the receptive field of the detecting mechanism at something less than two full cycles (e.g., see Meese, 2010; Meese & Summers, 2012). If, on the other hand, the stimuli were detected by mechanisms whose spatial extent matched the targets, then the summation slope would be −1/2, consistent with this ideal strategy (and the initial slopes of our own summation functions). So why might Wong and Levi's summation slope have been so steep? One possible factor is that, unlike in our study, Wong and Levi provided no indication of the precise location and size of the target stimulus. Thus, for their smallest stimulus sizes, we might expect a drop in their measured sensitivity (and therefore a steeper summation function) owing to (spatial) uncertainty (Pelli, 1985; Meese & Summers, 2012). But this still leaves the question of why the summation functions measured by us should be so much shallower for CM than for LM. At first glance, it might seem that summation for CM is not as extensive as it is for LM, but if that had been so, then we would not have found the strong benefit of filling in the holes in the Swiss cheese stimulus. Similarly, although the shallow summation functions might be explained by a rapidly accelerating contrast response exponent for CM ($p \gg 2$), that would not be consistent with the results and modeling in Experiment 2. As we mentioned in the Results and discussion section, a more likely account of the shallow CM functions relates to the decline in sensitivity to our carrier away from the center of the display. We used high-pass noise to avoid within channel effects of noise masking of the target from the carrier. However, by doing this, sensitivity to our carrier would have declined quite rapidly with eccentricity (Baldwin et al., 2012), attenuating the effective depth of signal modulation in the periphery, reducing sensitivity, and thereby the summation slope. Furthermore, filling in the holes of the Swiss cheese would not suffer from this problem because that operation does not involve further encroachment of peripheral retina (note that the check modulations were in sine-phase with the center of the display). Nonetheless, regardless of these details and the differences in stimulus design, it is gratifying that one of the main conclusions of Wong and Levi (2005) is the same as one of ours: that second-order (CM) signals are summed over multiple signal cycles, similar to first-order signals (LM).

Other factors

The cue-invariant spatial summation found here is closely related to the eye-invariant spatial summation found at detection threshold (Meese & Summers, 2009) and above (Meese & Baker, 2011; i.e., binocular summation for interdigitated signals across the eyes is very similar to that found for binocular superposition). Similarly, other studies suggest pooling over time and orientation (Meese, 2010; Baker, Meese, & Georgeson, 2013; Meese & Baker, 2013). Thus, the emerging picture is that early stages of vision analyze the image in terms of its basic properties, and that later stages pool across those dimensions. However, the details of how that pooling is controlled remain to be elaborated, and some dimensions are clearly handled differently from others. For example, other than at very low spatial frequencies (Meese & Baker, 2011; Gheiratmand, Meese, & Mullen, 2013), it seems that pooling across the eye is mandatory. In contrast, although the study here shows that spatial pooling does take place across LM and CM, separate LM and CM pathways presumably remain available to the observer, since LM and CM stimuli are not easily confused. Furthermore, separate pathways would also be needed for higher level analysis in which LM and CM information is used to determine whether luminance changes in the retinal image derive from changes in illumination or changes in material (Schofield, Hesse, Rock, & Georgeson, 2006). Similarly, in Experiment 2 we found very little effect of signal phase on summation ratios and, for simplicity, we dispensed with the phase term in our modeling. However, this does not necessarily mean that the spatial integrator that we have identified pools indiscriminately over phase; that possibility has not been tested directly. Indeed, since the phase relationship between LM and CM is valuable for determining shape from shading (Schofield, Rock, & Georgeson, 2010) this information must be preserved within the relevant visual module.

We are left with one final puzzle, however. The scheme in Figure 5 provides a good account of all of our results here, and involves probability summation across orientation channels. The cross-oriented stimuli here were always LM-CM pairings, but since the co-oriented area summation mechanisms pool over LM and CM, then the scheme here makes the same predictions for cross-oriented LM-LM (and CM-CM) pairings. However, for cross-oriented interdigitated Battenberg elements, Meese (2010) found summation ratios closer to 1.5 dB, implying a Minkowski exponent

of about 2 across orthogonal orientation channels (following earlier square-law transduction of signal contrasts), consistent with a second stage of square-law transduction and linear summation across orientation. Whether these different conclusions about the final summation stage owe to subtle differences in the stimulus type, a model detail relating to LM-CM pairings that we have overlooked, or something else, remains unclear. However, we do note the following possibility. If there were three different classes of co-oriented area pooling mechanisms: pure LM, pure CM, and LM-CM mixes, then the inconsistency above could be resolved by supposing linear summation across only the orthogonal pure mechanisms of the same type (i.e., either both LM or both CM), leaving probability summation across all other combinations.

Keywords: area summation, second order, cue combination, computational modeling

Acknowledgments

This work was supported by an Engineering and Physical Sciences Research Council (EPSRC) UK grant to TSM and Mark Georgeson (EP/H000038/1). We would like to thank Mark Georgeson for code to generate the stimuli and for helpful discussions regarding the modeling. We also thank Nicolaas Prins and an anonymous reviewer for helpful comments on an earlier version of this manuscript. Some of this work was presented at the European Conference of Visual Perception, 2012 in Alghero, Sardinia. This work was performed under the guidelines of the original statement of Helsinki.

Commercial relationships: none.
Corresponding author: Tim S. Meese.
Email: t.s.meese@aston.ac.uk.
Address: Centre for Vision and Hearing Research, School of Life and Health Sciences, Aston University, Birmingham, UK.

References

Allard, R., & Faubert, J. (2007). Double dissociation between first-and second-order processing. *Vision Research, 47,* 1129–1141. doi:10.1016/j.visres.2007.01.010. [PubMed]

Allen, H. A., Hess, R. F., Mansouri, B., & Dakin, S. C. (2003). Integration of first- and second-order orientation. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 20*(6), 974. doi:10.1364/JOSAA.20.000974. [PubMed]

Baker, D. H., & Meese, T. S. (2011). Contrast integration over area is extensive: A three-stage model of spatial summation. *Journal of Vision, 11*(14):14, 1–16, http://www.journalofvision.org/content/11/14/14, doi:10.1167/11.14.14. [PubMed] [Article]

Baker, D. H., & Meese, T. S. (2014). Measuring the spatial etent of texture pooling using reverse correlation. *Vision Research, 97*, 52–58, doi:10.1016/j.visres.2014.02.004.

Baker, D. H., & Meese, T. S. (2012). Zero-dimensional noise: The best mask you never saw. *Journal of Vision, 12*(10):20, 1–12, http://www.journalofvision.org/content/12/10/20, doi:10.1167/12.10.20. [PubMed] [Article]

Baker, D. H., Meese, T. S., & Georgeson, M. A. (2013). Paradoxical psychometric functions ("swan functions") are explained by dilution masking in four stimulus dimensions. *I-Perception, 4*(1), 17–35. doi:10.1068/i0552. [PubMed]

Baldwin, A. S., & Meese, T. S. (2015). Fourth-root summation of contrast over area: No end in sight when sensitivity is compensated by a witch's hat. In preparation.

Baldwin, A. S., Meese, T. S., & Baker, D. H. (2012). The attenuation surface for contrast sensitivity has the form of a witch's hat within the central visual field. *Journal of Vision, 12*(11):*23*, 1–17, http://www.journalofvision.org/content/12/11/23, doi:10.1167/12.11.23. [PubMed] [Article]

Berns, R. S., Motta, R. J., & Gorzynski, M. E. (1993). CRT colorimetry. Part I: Theory and practice. *Color Research and Application, 18*(5), 299–314. doi:10.1002/col.5080180504.

Bradley, C., Abrams, J., & Geisler, W. (2014). Retina-V1 model of detectability across the visual field. *Journal of Vision, 14*(12):22, 1–22, http://www.journalofvision.org/content/14/12/22, doi:10.1167/14.12.22. [PubMed] [Article]

Chubb, C., & Sperling, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A, 5*(11), 1986–2007. doi:10.1364/JOSAA.5.001986. [PubMed]

Dakin, S. C., & Mareschal, I. (2000). Sensitivity to contrast modulation depends on carrier spatial frequency and orientation. *Vision Research, 40*(3), 311–329, doi:10.1016/S0042-6989(99)00179-0.

Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision.

Vision Research, 21, 1041–1053. doi:10.1016/0042-6989(81)90009-2. [PubMed]

Georgeson, M. A., & Schofield, A. J. (2002). Shading and texture: Separate information channels with a common adaptation mechanism? *Spatial Vision, 16*(1), 59–76. doi:10.1163/15685680260433913. [PubMed]

Gheiratmand, M., Meese, T. S., & Mullen, K. (2013). Blobs versus bars: Psychophysical evidence supports two types of orientation response in human color vision. *Journal of Vision, 13*(1):2, 1–13, http://www.journalofvision.org/content/13/1/2, doi:10.1167/13.1.2. [PubMed] [Article]

Hess, R. F., Baker, D. H., May, K., & Wang, J. (2008). On the decline of 1st and 2nd order sensitivity with eccentricity. *Journal of Vision, 8*(1):19, 1–12, http://www.journalofvision.org/content/8/1/19, doi:10.1167/8.1.19. [PubMed] [Article]

Johnson, A. P., & Baker, C. L. Jr., (2004). First-and second-order information in natural images: a filter-based approach to image statistics. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 21*(6), 913. doi:10.1364/JOSAA.21.000913. [PubMed]

Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research, 31,* 679–691. doi:10.1016/0042-6989(91)90009-T. [PubMed]

Landy, M. S., & Oruç, İ. (2002). Properties of second-order spatial frequency channels. *Vision Research, 42*(19), 2311–2329. doi:10.1016/S0042-6989(02)00193-1. [PubMed]

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70,* 1458–1471. doi:10.1364/JOSA.70.001458. [PubMed]

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 7,* 923–932. doi:10.1364/JOSAA.7.000923. [PubMed]

Meese, T. S. (2004). Area summation and masking. *Journal of Vision, 4*(10):8, 930–943, http://www.journalofvision.org/content/4/10/8, doi:10.1167/4.10.8. [PubMed] [Article]

Meese, T. S. (2010). Spatially extensive summation of contrast energy is revealed by contrast detection of micro-pattern textures. *Journal of Vision, 10*(8):14, 1–21, http://www.journalofvision.org/content/10/8/14, doi:10.1167/10.8.14. [PubMed] [Article]

Meese, T. S., & Baker, D. H. (2011). Contrast summation across eyes and space is revealed along the entire dipper function by a "Swiss cheese" stimulus. *Journal of Vision, 11*(1):23, 1–23, http://www.journalofvision.org/content/11/1/23, doi:10.1167/11.1.23. [PubMed] [Article]

Meese, T. S., & Baker, D. H. (2013). A common rule for integration and suppression of luminance contrast across eyes, space, time, and pattern. *I-Perception, 4*(1), 1–16. doi:10.1068/i0556. [PubMed]

Meese, T. S., & Hess, R. F. (2007). Anisotropy for spatial summation of elongated patches of grating: A tale of two tails. *Vision Research, 47,* 1880–1892. doi:10.1016/j.visres.2007.04.008. [PubMed]

Meese, T. S., & Summers, R. J. (2007). Area summation in human vision at and above detection threshold. *Proceedings of the Royal Society B: Biological Sciences, 274*(1627), 2891–2900. doi:10.1098/rspb.2007.0957. [PubMed]

Meese, T. S., & Summers, R. J. (2009). Neuronal convergence in early contrast vision: Binocular summation is followed by response nonlinearity and area summation. *Journal of Vision, 9*(4):7, 1–16, http://www.journalofvision.org/content/9/4/7, doi:10.1167/9.4.7. [PubMed] [Article]

Meese, T. S., & Summers, R. J. (2012). Theory and data for area summation of contrast with and without uncertainty: Evidence for a noisy energy model. *Journal of Vision, 12*(11):9, 1–28, http://www.journalofvision.org/content/12/11/9, doi:10.1167/12.11.9. [PubMed] [Article]

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 2,* 1508–1532. doi:10.1364/JOSAA.2.001508. [PubMed]

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research, 21,* 409–418. doi:10.1016/0042-6989(81)90169-3. [PubMed]

Schofield, A. J. (2000). What does second-order vision see in an image? *Perception, 29*(9), 1071–1086. doi:10.1068/p2913. [PubMed]

Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: Separate mechanisms with similar behaviour. *Vision Research, 39,* 2697–2716. doi:i:10.1016/S0042-6989(98)00284-3. [PubMed]

Schofield, A. J., Hesse, G., Rock, P. B., & Georgeson, M. A. (2006). Local luminance amplitude modulates the interpretation of shape-from-shading in textured surfaces. *Vision Research, 46*(20), 3462–3482. doi:10.1016/j.visres.2006.03.014. [PubMed]

Schofield, A. J., Rock, P. B., & Georgeson, M. A. (2010). What is second-order vision for? Discrim-

inating illumination versus material changes. *Journal of Vision, 10*(9):2, 1–18, http://www.journalofvision.org/content/10/9/2, doi:10.1167/10.9.2. [PubMed] [Article]

Smith, A. T., & Ledgeway, T. (1997). Separate detection of moving luminance and contrast modulations: Fact or artifact? *Vision Research, 37*(1), 45–62. doi:10.1016/S0042-6989(96)00147-2. [PubMed]

Sukumar, S., & Waugh, S. J. (2007). Separate first-and second-order processing is supported by spatial summation estimates at the fovea and eccentrically. *Vision Research, 47,* 581–596. doi:10.1016/j.visres.2006.10.004. [PubMed]

Summers, R. J., & Meese, T. S. (2009). The influence of fixation points on contrast detection and discrimination of patches of grating: Masking and facilitation. *Vision Research, 49*(14), 1894–1900. doi:10.1016/j.visres.2009.04.027. [PubMed]

Tyler, C. W., & Chen, C. C. (2000). Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation. *Vision Research, 40,* 3121–3144. doi:10.1016/S0042-6989(00)00157-7. [PubMed]

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *The British Journal of Mathematical and Statistical Psychology, 18,* 1–10. doi:10.1111/j.2044-8317.1965.tb00689.x. [PubMed]

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313. doi:10.3758/BF03194544. [PubMed]

Wilson, H., Ferrera, V., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience, 9*(1), 79–97. doi:10.1017/S0952523800006386. [PubMed]

Wong, E. H., & Levi, D. M. (2005). Second-order spatial summation in amblyopia. *Vision Research, 45,* 2799–2809. doi:10.1016/j.visres.2005.05.020. [PubMed]